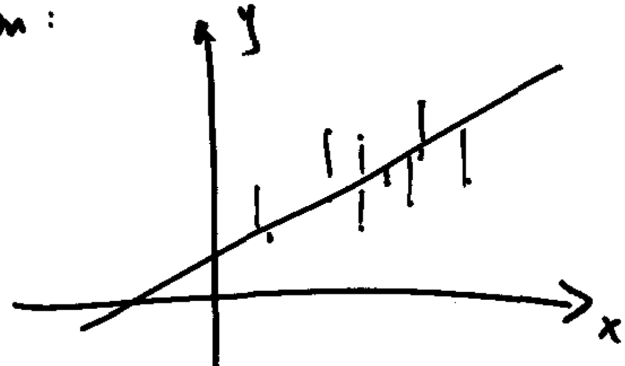
Stat 564
9-26-17

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

1

Simple linear regression:

$$y = \beta_0 + \beta_1 x + \varepsilon$$



Goal: minimize the sum of the squares of the residuals

Assume we have n observations

2

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_1 + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_2 + \varepsilon_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_n + \varepsilon_n \end{cases}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}_{n\times 2} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}_{2\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

③

$$\underset{n\times 1}{Y} = \underset{n\times 2}{X}\,\underset{2\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$$

$$SSE = \sum_{i=1}^{n}\left[y_i - (\beta_0 + \beta_1 x_i)\right]^2$$

"sum of squares due to error"

$$SSE = \sum_{i=1}^{n}\left[y_i - \beta_0 - \beta_1 x_i\right]^2$$

$$\frac{\partial SSE}{\partial \beta_0} = \sum_{i=1}^{n} \cancel{2}[y_i - \beta_0 - \beta_1 x_i](\cancel{-1}) \overset{set}{=} 0$$

$$\sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum x_i = 0$$

④

$$n\beta_0 = \sum y_i - \beta_1 \sum x_i$$

$$\beta_0 = \frac{1}{n}\sum y_i - \beta_1 \frac{1}{n}\sum x_i$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\frac{\partial SSE}{\partial \beta_1} = \sum_{i=1}^{n} \cancel{2}[y_i - \beta_0 - \beta_1 x_i](\cancel{-}x_i) \overset{set}{=} 0$$

$$\sum_{i=1}^{n} y_i x_i - \beta_0 \sum x_i - \beta_1 \sum x_i^2 = 0$$

$$\sum x_i y_i - (\bar{y} - \hat{\beta}_1 \bar{x})\sum x_i - \hat{\beta}_1 \sum x_i^2 = 0$$

(5)

$$\sum x_i y_i - \bar{y}\sum x_i = \hat{\beta}_1 \sum x_i^2 - \hat{\beta}_1 \bar{x} \sum x_i$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \bar{y}\sum x_i}{\sum x_i^2 - \bar{x}\sum x_i}$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{S_{xy}}{S_{xx}}$$

(6)

Summary:

$$\boxed{\begin{array}{l} \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \end{array}}$$

slope
intercept
of the least squares line

Our prediction equation is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Note: see what happens if $x = \bar{x}$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

7

$$\hat{y} = (\bar{y} - \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 \bar{x}$$
$$= \bar{y}$$

So the least square line always passes through $(\bar{x}, \bar{y})$, which is the center of mass of the observations

---

Alternate forms of S_{xy}:

$$S_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

8

$$= \sum x_i y_i - n \frac{(\sum x_i)(\sum y_i)}{n^2}$$
$$= \sum x_i y_i - n \bar{x} \bar{y}$$

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum (x_i y_i - \bar{x} y_i - x_i \bar{y} + \bar{x}\bar{y})$$
$$= \sum x_i y_i - \bar{x} \underbrace{\sum y_i}_{n\bar{y}} - \bar{y} \underbrace{\sum x_i}_{n\bar{x}} + n \bar{x} \bar{y}$$
$$= \sum x_i y_i - n \bar{x} \bar{y} = S_{xy}$$

(9)

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$= \sum (x_i - \bar{x}) y_i - \underbrace{\sum (x_i - \bar{x}) \bar{y}}_{\bar{y} \underbrace{\sum (x_i - \bar{x})}_{\underbrace{\sum x_i - n\bar{x}}_{0}}}$$

$$\boxed{\begin{aligned} S_{xy} &= \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} \\ &= \sum x_i y_i - n \bar{x} \bar{y} \\ &= \sum (x_i - \bar{x})(y_i - \bar{y}) \\ &= \sum (x_i - \bar{x}) y_i \\ &= \sum x_i (y_i - \bar{y}) \end{aligned}}$$

(10)

Similarly,

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$= \sum x_i^2 - n \bar{x}^2$$

$$= \sum (x_i - \bar{x})^2 = (n-1) s_x^2$$

$$= \sum x_i (x_i - \bar{x})$$